

1970

Patterns of change in academic performance in college and their correlates with pre-college variables

John Edward Klingensmith
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>



Part of the [Psychology Commons](#)

Recommended Citation

Klingensmith, John Edward, "Patterns of change in academic performance in college and their correlates with pre-college variables " (1970). *Retrospective Theses and Dissertations*. 4332.
<https://lib.dr.iastate.edu/rtd/4332>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

71-7293

KLINGENSMITH, John Edward, 1923-
PATTERNS OF CHANGE IN ACADEMIC PERFORMANCE IN
COLLEGE AND THEIR CORRELATES WITH PRE-COLLEGE
VARIABLES.

Iowa State University, Ph.D., 1970
Psychology, general

University Microfilms, Inc., Ann Arbor, Michigan

PATTERNS OF CHANGE IN ACADEMIC PERFORMANCE IN COLLEGE
AND THEIR CORRELATES WITH PRE-COLLEGE VARIABLES

by

John Edward Klingensmith

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Psychology

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Ames, Iowa

1970

TABLE OF CONTENTS

	Page
INTRODUCTION	1
STATEMENT OF PROBLEM	21
METHOD	24
STUDY 1: RESULTS AND DISCUSSION	38
STUDY 2: RESULTS AND DISCUSSION	49
SUMMARY	74
REFERENCES	76
ACKNOWLEDGMENTS	80
APPENDIX A: FRESHMAN PROGRAM FOR ENGINEERING STUDENTS	81
APPENDIX B: FORTRAN PROGRAM FOR COMPUTING ITEM- CRITERION CORRELATIONS	84
APPENDIX C: SUPPLEMENTARY TABLES	88

INTRODUCTION

That human behavior is ultimately predictable and hence controllable has been the faith and the hope of modern scientific psychology. In the particular area of educational psychology, the current unrest on American campuses and elsewhere has occasioned some serious reconsideration of the validity of this faith. For one thing, it is all too apparent that psychologists failed to foresee the present disillusion with traditional educational goals and practices. For another, there is little evidence that psychologists or educators fully understand the problem even now or know what exactly to do about it.

There is evidence, however, that psychologists are taking a new look at some of their cherished assumptions. After a perhaps excessive flirtation with various forms of behaviorism, they have begun once more to take the organism seriously. But the course of this change in viewpoint is not going to be a smooth one. The recent discussion by Jensen (1969) of studies regarding the effect of genetic factors on measured IQ has stirred up the old nature-nurture controversy. It is unfortunate that much of the merit of these studies has been overlooked in the storm of political controversy which they have generated. Still, they have

served to remind us of the wide split which exists in many behavioral disciplines between those who would look for the explanation of behavior in the nature of the organism and those who would look for it in the characteristics of the environment. While many researchers try to take both poles into account, most of them seem to end up nearer one end of the continuum than the other.

Trait-psychometric psychology, which is at least as behavioristic as it is organismic, has simply not provided an adequate model for the discussion of human behavior in formal educational contexts. There is, consequently, a search underway for a broader conceptualization. Both educational psychologists and personality theorists seem to be moving toward an interactionist stance (see, e.g., Cronbach, 1957; Mischel, 1969); that is, a standpoint in which individual traits and abilities, the characteristics of the environmental context, and the interaction of these two complexes are all taken into account.

The immediate concern of the present study is to take a new look at the old problem of the prediction of academic success in college. After a brief review of the history of, and modern developments in, the prediction of academic success, a method for extracting more and different kinds of information from traditional types of data will be presented

and discussed. It would be foolish to suggest that current selection and admission procedures are a major cause of student unrest--though equality of educational opportunity for the underprivileged is involved--or to suggest that improvements in these procedures would solve all problems on our campuses. The question is one rather of finding ways to utilize educational resources more effectively and efficiently in a time when our basic institutions are accused of failing to serve the needs of people. It is hoped that the present effort will contribute in some way to the discarding of old molds and to encouraging the search for new ways of thinking about a perennial educational concern.

The Prediction of Academic Success: Historical Perspective

Few problems in education have received more research attention than has that of the prediction of academic success, particularly in college. Garrett (1949) reviewed the literature for the decade of the 1930's and 40's and found 194 studies which specifically treated various aspects of this problem. Ten years later, Fishman and Pasanella (1960) found that 580 additional articles on the subject had been produced in the decade of the 50's. In the past decade there has been no diminution of interest in this problem although,

as will be indicated, there has been some shift of research emphasis.

Intellective predictors

Most of the research referred to above seems to have been a natural by-product of the mass testing movement in America. As the demand for college education increased and it became necessary to allocate resources more wisely, administrators sought whatever help they could find in estimating the probable success of those admitted to college. It was a fairly easy matter to use the results of efforts to validate the mass testing instruments as a basis for predicting academic performance. Hence prediction developed and was largely applied in an admissions context, mostly in coed and state-supported schools. While such predictions were far from perfect, they did and still do enjoy much success, and the techniques they use have been widely adapted in counseling and placement services.

The usual approach in this type of prediction relies on correlation and regression techniques. Predictors--typically, college entrance exams and measures of previous academic performance--are correlated with criteria--typically, first-term grade point average (GPA). A small number of "best" predictors are decided upon and are combined in a multiple regression equation. While there have always been those who

objected to the impersonality of this technique, in an important book Meehl (1954) showed that a multiple correlation approach with two or three objective measures generally yields higher predictive values for this type of criterion than do subjective evaluations by trained personnel using the same data plus an interview.

Despite their obvious validity and usefulness, such prediction methods were not without their shortcomings. Of the 580 studies noted by Fishman and Pasanella, nearly 70 percent involved only the correlation of intellectual measures with intellectual criteria. Many of them recognized the enormous drop in the value of the multiple-correlation if any kind of preselection resulted in a restricted range of talent among the testees. And even with the best prediction equations using intellectual factors, the resulting R^2 's were disappointingly small. These limitations encouraged an expansion of interest in the search for other, perhaps non-intellectual, predictors.

Nonintellectual predictors

At the time of their review, Fishman and Pasanella noted an increasing number of studies devoted to the search for nonintellectual predictors; but there were few reports of success in using nonintellectual predictors, either alone or in conjunction with intellectual predictors. The

nonintellective predictors which received most research attention were personality tests such as the California Psychological Inventory, projective instruments such as the Rorschach, and interest inventories such as the Strong Vocational Interest Blank.

In a later review of such efforts, Mayhew (1965, p. 40) came to the following conclusion:

"Adding a measure of personality to previously employed high school rank and academic aptitude measures has sometimes increased predictive powers slightly, but such an addition has also resulted in actual decreases in predictive success."

Lavin (1965) came to a similar conclusion: background characteristics add little to prediction once grades and a measure of general intellectual ability are already in the prediction equation; their effectiveness, if any, would seem to be in functioning as moderator variables. One of the points of the present research is to suggest a different way to extract information from nonintellective measures in view of the prediction problem.

Differential prediction

In a somewhat different direction, Horst (1957) became a strong advocate of differential prediction. Considering college GPA as too global a criterion, he sought to develop techniques for working in terms of grades assigned for particular courses or in particular skills. While this

approach showed much early promise, it proved to be very expensive in terms of money and of requirements for professional competence to apply it. Also, and perhaps most important today, it requires too frequent a revision in a time of rapid curricular change.

Increased statistical sophistication

Researchers have displayed much ingenuity in trying to come up with a combination of approaches and techniques which might provide better prediction. Burket (1964) found that, in using multiple predictors for one criterion, factor weights of the predictors showed greater sample-to-sample stability than did the predictors themselves. Hence he proposed reducing a large pool of predictors by factoring prior to prediction selection and weighting. Lunneborg, in a series of studies (1966, 1968a, 1968b), investigated the differences between differential and absolute prediction as related to the use of factor scores and biographic variables. Regarding factor analysis of predictors, he found factor scores good for absolute prediction but he also found the original predictors much superior to the factored ones for differential prediction. He likewise found biographic data much more useful in the differential prediction of academic achievement, and confirmed once more the common

finding that absolute prediction relies most heavily on measures of prior academic performance.

Other researchers tried to develop more sophisticated statistical and psychometric techniques for application to the prediction problem. Brown and Scott (1966) investigated various moderator variables but found none that resulted in any significant increase in predictive validity. Klein, Rock and Evans (1968), in an investigation of the use of multiple moderators in the prediction equation, suggested the notion of working backwards from academic performance to a search for precollege correlates of performance. Ohnmacht (1968) explored the predictability of change in academic achievement defined as the regressed estimate of true change; i.e., he corrected change estimates for reliability of both initial and final measures. All of these notions will find application and extension in the research reported in this study.

Related research

The vast body of experimental research on learning, though it is potentially important to educational psychology, has not been mentioned as pertinent to the present study. The long series of studies by Fleishman, on the general theme of changes in the factorial content of motor skills as a function of practice, does however deserve mention. In the

latest of these studies, Fleishman and Ellison (1969) investigated the relation between personality inventories and certain perceptual-motor tasks. They found, once again, that ability measures predicted performance level but that personality measures did not. They also found that learning rate was not predictable from either personality or ability measures.

One of the problems which has constantly plagued prediction research is the nature of the single criterion which is most often used as the measure of academic achievement, the grade point average. Juola (1968), in a summary of the difficulties which attend the use of this criterion, has shown that, while there is evidence that students at Michigan State University have gotten brighter (as measured by the College Qualification Test), the all-freshmen GPA has tended to remain constant. He calls this a "nonyielding grading distribution" and says that it partly accounts for the attrition rate in certain colleges. He also found definite evidence of teacher bias in grading not related to varying levels of ability in students. Some research by the present author on grade distributions in multi-section courses at Iowa State University demonstrated this same phenomenon (Menne and Klingensmith, 1968).

Conclusion

Today, then, academic prediction in the traditional form has little to contribute to the solution of the monumental problems of understanding the behaviors of college students. What seems to have happened is that a useful technique for validating measures came to be routinized in the service of admission and selection procedures. The prediction of academic success tended to become frozen and even semi-sacred at the very time that pressures were mounting for a reconsideration of the goals and the criteria on which the whole procedure rested.

Toward an Interactionist Conceptualization

As was suggested in the foregoing discussion, disillusionment with traditional prediction results is nothing very new. But only in the past decade or so have alternatives begun to be proposed and investigated. At the risk of some oversimplification, these efforts may be grouped into two major categories: studies of aptitude-treatment interaction, and school-effects studies. Another potentially important category is represented in Flanders' verbal interaction analysis, but this latter deals with the social climate of the classroom and does not appear to be directly related to the present research.

Aptitude-treatment interaction

The leading exponent of this position is Lee Cronbach. In an important early paper, Cronbach (1957, p. 680) stated very clearly the interactionist faith and position:

"Applied psychologists should deal with treatments and persons simultaneously. Treatments are characterized by many dimensions; so are persons. The two sets of dimensions together determine a payoff surface. For any practical problem, there is some best group of treatments to use and some best allocation of persons to treatments. We can expect some attributes of persons to have strong interactions with treatment variables. These attributes have far greater practical importance than the attributes which have little or no interaction."

Following his own advice, Cronbach has been a leader in the effort to design research which might reveal the complex relations of instructional methods to pupil aptitudes.

In a lengthy report, Cronbach and Snow (1969) reviewed the efforts which have been made in the investigation of these relationships, as well as the results of their own research. The progress to date has not been impressive. Very few aptitude-treatment interactions have been identified, and where they have been, there is little evidence to support their stability. The authors suggest as the possible source of this difficulty the fact that the treatments used in such experimentation are usually too short and too artificial. One line of attack on this difficulty is to gather large amounts of longitudinal data on ongoing, innovative educational programs. These latter, however, are not easy to

come by. A second line is to produce more adequately designed experimental studies. A third possibility, which they do not consider, is to seek to identify, from existing or traditional types of data, those variables pertinent to the interaction process. It is an effort in the latter direction with which the present research is concerned.

The kind of research which Cronbach advocates is admittedly high-risk research; i.e., it may cost a lot and produce very little. And Bracht (1969) has pointed out that it involves an essential dilemma. Any aptitude-method interaction which might be discovered would have to have a high degree of confirmation in order to be politically acceptable and implementable in educational practice. In order for such research to be feasible, however, the researcher must necessarily deal with broadly defined instructional treatments and student types. Yet Bracht presents evidence that the search for interactions using broadly defined treatments is probably going to remain fruitless.

In addition to the already abundant complexities of aptitude-treatment interaction research, Mitchell (1969) has suggested that it may have to deal with even greater complexity. The interactions, if they exist, may have a developmental aspect such that time becomes the third factor in a three-way interaction. In other words, interactions may appear and disappear developmentally. And finally, it

may be necessary to define different payoffs in terms of varying educational objectives. In that case, one again ends up with three-way interactions involving aptitude, method and performance.

While this line of investigation is still in its infancy, and results so far have been less than spectacular, it has many implications for educational psychology. Should it be possible to find a more adequate match between student abilities and instructional techniques, many dissatisfactions with the college experience would probably disappear.

School-effects studies

While both the previous and present categories fall under the rubric of person-environment interaction (PEI), they differ principally in the manner in which they conceptualize the environment. In aptitude-treatment interactions, the environment is viewed in the rather limited terms of the instructional stimulus to which the student is exposed. In the present category, environment is much more broadly defined as the total school experience.

Two principal types of instruments have been utilized in the effort to make the kind of situational analysis which is the concern here. The first of these is represented by the College Characteristics Index (Stern, 1963) and by the College and Universities Environment Scales (Pace and Stern,

1958). These instruments are designed to get at what H. A. Murray called "beta press", the environment as perceived and/or interpreted by the observer, in this case the student himself. It is argued that for some purposes the college environment is what the students perceive it to be.

The second type of instrument is represented by the Environmental Assessment Technique (Astin and Holland, 1961). This is designed to measure what Murray called "alpha press", the environment as it exists in objective reality.

Both of these types are what must be called group-measuring instruments. While the extensive studies of Astin have shown that they can be quite successfully used to identify different types of college environments (see, e.g., Astin, 1962), they have a serious drawback when used for PEI analysis: they tend to confound person variance with environment variance. As Mitchell (1969) points out, this may be entirely acceptable from a phenomenological standpoint, but it is unacceptable in PEI analysis. The variance of student input characteristics should be taken into account before student behaviors are attributed to college press. This is the problem of "partialling out the variance" in school-effects studies, and it has given rise to some controversy between Astin and his critics (see, e.g., Werts and Watley, 1969; Astin, 1969). As yet, the difficult methodological problems

arising in this kind of research have not been satisfactorily solved.

At best, school-effects studies would seem to have demonstrated relatively small second-order effects. The task of discovering and defining the first-order effects, those environmental variables which affect the student directly, remains to be done. Nor have researchers yet solved such problems as the development of adequate theoretical rationales for the interpretation of person-environment interactions. It would be logical, for instance, since the instruments used are based upon Murray's press theory, to try to measure student "needs" in the same theoretical framework and then to seek a need-press fit. Stern (1963) made such an effort, but without impressive results; he found it impossible either to define the optimal press for a particular need or to discover appropriate norms for judging the need-press fit.

As Chickering, McDowell and Campagna (1969) observed, in a recent attempt to relate institutional differences to student behavior, American higher education has long proceeded on the twin assumptions that significant development occurs in the college years, and that the college environment is a major factor influencing that development. The research of Sanford, Katz, Newcomb, and others has uncovered abundant evidence that the first of these assumptions is

indeed true, at least in the area of personality and attitude development. But the systematic relationships postulated by the second assumption have as yet received little empirical support. Psychologists and educators are quite sure that the college environment must be an important factor in intellectual as well as personality development, but they have yet to point out the what, when and how of the process.

A Developmental Approach to the Prediction of Academic Performance

If prediction research is to serve a useful function today, it must move beyond its traditional limitations. One direction it might take is to look at the prediction problem from an interactionist standpoint.

Nearly all of the prediction work reviewed above has had two characteristics: it has used grade point average as the criterion variable, and it has been concerned with first-term or at most first-year college performance. The GPA has usually been the criterion for prediction principally because no one has yet proposed an acceptable alternative measure of "success" in college. Differential prediction has sought to use other criteria, but these often turn out to be either difficult to define or expensive to measure.

In the present study, the GPA is assumed to be a valid criterion of academic performance.

However, it should be not only possible but desirable to look beyond academic performance in the first term or first year. Some steps have already been taken in this direction. Juola (1963) investigated the long-range prediction possibilities of college aptitude measures. He found a linearly decreasing correlation between these predictor scores and later term GPA, which he attributed to a restriction in range of abilities as attrition takes its toll over the college experience. In a later study (1966), he found that the best predictor of GPA for a given term was the cumulative GPA of the immediately preceding term and that pre-college ability measures added little to this predictor. He suggested that more effort be put into analyzing previous college achievement, and that pre-college ability measures might be of value only if the student were to enter a new field.

In a large-scale study at the University of Illinois, Humphreys (1968) also found that correlation between freshman predictors and semester GPA declined steadily from the freshman through the senior year, with common variance no more than eight percent in the final semester. From this he concluded that people change over the college experience and that aptitude for college work is far from being a

stable characteristic.¹ He then issued anew the plea that researchers undertake the major task of finding good non-academic predictors of GPA.

But it is possible to draw a different conclusion from the work of Humphreys and Juola. Drawing together some of the ideas touched upon earlier, one can postulate the existence of patterns of change in academic performance over the college experience, since it is not likely that the changes observed are merely random. Should such patterns emerge, it might be possible to relate pre-college measures of background, ability and personality to probable types of performance in college. Such relationships could then be used to gain insight into the nature of the interactions between person and college environment. While one would use the GPA as the principal measure in such an investigation, the concern would not be with the average itself but with the pattern in which it occurred.

The general model proposed by Schaie (1965) for the study of developmental problems provides a theoretical rationale for an investigation of this type. Schaie adopts Kessen's dictum that "a characteristic is developmental if

¹Humphreys does not seem to have considered the possibility of other explanations. For example, perhaps it is not the people who are changing so much as what is required of them later on in college.

if can be related to age in an orderly or lawful way" (Kessen, 1960). He then proceeds to define three additional terms:

age: the number of time units elapsed between birth and the point in time at which a response is recorded;

cohort: the total population of organisms born at the same time;

time of measurement: an index of the total environmental impact as occurring at a given temporal point.

Using these concepts, he offers the general model:

$$\text{response} = \underline{f}(\text{age}, \text{cohort}, \text{time}).$$

The three variables on the right hand side of this equation are not seen as independent but as reacting with one another.

In the present study it is not the number of years lived which is of primary concern but number of terms of college experience. Nor is it possible to give a precise operational definition of the term "cohort" in the proposed functional relationship. Nevertheless, it will be useful to have this model in mind in interpreting the results of this research.

A Concluding Perspective

A decade ago Webster, Freedman and Heist (see Sanford, 1962, p. 818) made the following observation:

"The problem of understanding the relationships between the process of acquiring knowledge, as measured by achievement tests, and measured abilities and other personality characteristics of students is a complex one, the investigation of which has hardly begun."

The problem is, if anything, even more important today, and there is little evidence that we have noticeably advanced toward its solution. The present research is intended as a step in that direction.

The notion of studying change in academic performance is not a new one, yet it has not been very thoroughly investigated in the present context. As Jensen (1969, p. 23) has observed, such studies may be the key to the solution of some of our most pressing educational problems:

"Perhaps our greatest hope of achieving equality of educational opportunity lies in the possibility of finding significant patterns of individual differences in the development of abilities and in taking advantage of these differences to create optimal instruction pupil interaction." (*italics added*)

STATEMENT OF PROBLEM

There are two focuses in this project. The first is a quasi-experimental study in which comparisons are made between a selected group of engineering students at Iowa State University who were subjected to an innovative engineering curriculum and a matched control group who followed the traditional program. The experimental treatment was to introduce courses in design at the beginning of the engineering curriculum of study rather than later on, as is usual. The experimental group participated in this innovative program on a voluntary basis. In order to ensure that any conclusions made in regard to the treatment objective were valid, a control group was carefully selected to match the experimental group on as many dimensions as possible.

The first phase of this study was directed toward testing Hypotheses I and II.

Hypothesis I: Engineering students who begin to study design principles early in their college career will obtain better term grades than do students who follow traditional programs.

Hypothesis II: Engineering students who begin to study design principles early in their college career will tend to change curricula and/or drop

out of school less frequently than do students who follow traditional programs.

The second, and major, focus of the present study was to investigate and describe the manner in which pre-college variables are related to various patterns of academic performance. A battery of ability, personality, and interest measures was administered to the subjects in both of the above groups at the beginning of their freshman year; the number and nature of these pretests, as well as their item and factorial content, are discussed at length in the next section of this paper. Thus, this phase of the study is correlational rather than experimental.

This second phase of the study dealt with Hypotheses III and IV.

Hypothesis III: Meaningful patterns of academic performance exist and are discernable in engineering students.

Hypothesis IV: These patterns of performance are related to, and can to some degree be predicted from, pre-college test variables.

The "patterns of performance" referred to in these hypotheses will be operationally defined in terms of trend lines. This definition and its rationale will be made explicit in the following section of this paper.

In order to provide a theoretical framework in which to discuss the results of this second phase of the investigation, use is made of the general developmental model proposed by Schaie and described in the previous section. According to this model, a response is a function of age, cohort and time. The full specification of this model would require more controls than were possible in the present study. What is being presented is only an approximation to the model. For the purpose of this research, age is measured in quarters of completed college work. While a summer in which some subjects took additional courses and others did not intervened between the first and last three quarters of the aging process, the age periods are considered as essentially contiguous and of the same duration. Cohort is the least precisely defined and measured construct in this study; it will consequently enter the discussion specifically only when comparisons are made between the performances of the studied group and their peers in the total college population at Iowa State University. Time is defined as those points at which final grades for a quarter are assigned and grade point average is computed.

METHOD

Setting

In the fall quarter of 1967, the Iowa State University College of Engineering introduced an experimental program in which a sequence of courses incorporating design principles was made a part of the first-year engineering curriculum. It had been theorized, after a series of discussions, that the high attrition rate in Engineering could be explained in part by the fact that in the traditional curriculum--in which students spend the first two or three years taking courses in mathematics and science before coming to design--many students simply lose interest or become discouraged. By introducing engineering design early in the course of studies, it was hoped that the students would become more involved and would have less tendency to drop out of engineering curricula. It was also hoped that this increased interest would result in better academic performance. (Background information and details on this experimental program will be found in Appendix A.)

The present study grew out of, and is an extension of, the evaluation phase of this experimental program.

Subjects

In order to test the effectiveness of this curricular innovation, it was decided to select a sample of entering male freshmen which would be proportionally representative of majors in the College of Engineering and invite them to participate on a voluntary basis in the new program. Once the experimental group had thus been formed, a matched control group was selected. The control group originally had the same representation as the experimental group and was made, on the basis of admissions information and orientation test data, to match the experimental group as exactly as possible. Table 1 shows the composition of the two groups. When data were assembled for the present study, it was found that one of the original control group subjects was female, several were upperclassmen, and some had failed to take the additional tests required. Elimination of these subjects from the study accounts for the fact that the N's in the two groups are not equal. It will be shown later, however, that the two groups were in fact almost identical on the variables of interest to this study.

Table 1. Composition of groups by major in the College of Engineering

Major	Experimental	Control
Agricultural Engineering	-	1
Aerospace Engineering	26	21
Ceramic Engineering	1	1
Chemical Engineering	11	11
Civil Engineering	17	18
Electrical Engineering	30	22
Engineering Operations	1	1
Engineering Science	3	3
Industrial Engineering	4	6
Mechanical Engineering	<u>22</u>	<u>18</u>
	115	102

Procedure

In 1967, all freshmen entering Iowa State University were required to take three tests during the summer orientation program: the Minnesota Scholastic Aptitude Test (MSAT); English Expression, Form 1c, of the Cooperative English Tests (ENGL); and the Iowa State Mathematics Placement Test (MATH). From admissions records were obtained the high school rank (HSR) and the composite score on the American College Testing Program (ACT) for the subjects in this study. These ACT scores were available for only those students who came from

high schools in which this test had been administered. For many of the latter students, there was also available the ACT 1968 Class Profile data (ACT-B).

To supplement this information, an effort was made to obtain scores on the following instruments for all the subjects in both groups: the Engineering Attitude Survey (EAS); the Allport-Vernon-Lindzey Study of Values (AVL); the Gough Adjective Check List (ACL); and the Strong Vocational Interest Blank for Men (SVIB). For the purpose of the present study, all of these are considered precollege "test scores".

It was originally intended to retest, at the end of the first year, all the subjects in both groups on the four last-mentioned tests. Because of certain changes in personnel, this part of the original project was not completed.

The principal aim of the present study is to examine the relationships which might exist between college predictors and certain criteria of college academic performance. It utilizes the data gathered in the original project, but it goes beyond the traditional prediction study in two ways: first it expands the concept of "predictor" to include each item on each of the pretests as well as certain additional scores derived from these items; second, it expands the concept of "criterion" by generating trend indicators from the quarter GPA's rather than looking at course grades. The first of

these notions will be described under independent variables and the second under dependent variables.

Independent variables

For each of the subjects in this study the three orientation tests (MSAT, ENGL, and MATH) were scored on an IBM 1230 Optical Reader and the resulting output was transformed to dichotomous (1 = right, 0 = wrong) item data on punched cards.

It was desired to use these data as exhaustively as possible. For the MSAT there are three logical subscales: Reading (33 items), Same-Opposites (15 items), and Analogies (30 items). For the ENGL there are two logical subscales: Effectiveness (30 items), and Mechanics (60 items). These five subscale scores were computed for each subject in the study.

In order to examine the possible factorial content of these three tests, the following procedure was used for each of them. A random sample of about 700 persons was taken from among all Iowa State freshmen who had taken the test in 1967, 1968 or 1969. This number provided the approximate 10-1 subject-to-item ratio recommended by Nunnally (1967) for a factor analysis. Dichotomous item scores were derived for each of these persons, as described above, and the resulting data were subjected to principal components factor

Table 2. Variance removed by item factor analyses of the MSAT, ENGL and MATH tests

Test	% of common variance removed by factor				% of total variance removed
	I	II	III	IV	
MSAT	36.75	16.86	18.92	27.47	17.29
ENGL	35.28	23.66	20.22	20.84	10.58
MATH	26.84	30.42	15.61	27.13	23.53

analysis using highest correlation for row and column in the diagonals. Independent random samples were used for each of the three tests. For each test eight factors were extracted and on the basis of eigenvalues four factors for each were retained. These were then subjected to Varimax rotation.

The percent of common variance removed by each factor and the percent of total variance accounted for by the four factors are given for each analysis in Table 2. Using the rotated factor loadings, four factor scales were determined for each of the tests, based on the following arbitrary criteria:

(a) an item may be included in only one scale; (b) the loading must have an absolute value of 0.75 or greater; and (c) a loading must exceed the loadings on other factors by 0.15.

The aim of these stringent requirements was to eliminate overlap and produce scales which were as "pure" as possible. On the basis of the factors thus derived, factor scale scores

were computed for each of the subjects in the present study. The nature of the resulting factor scales will be discussed in the following section of this paper.

The ACT-B data were obtained from the American College Testing Program in the Supplementary Punch Card Report. These data were already coded, with a few exceptions, in a format suitable for this study. Forty-six variables, tapping such information as Goals and Aspirations, Student Personnel Needs, Non-academic Achievements, College Attractions, and Demographic Data were selected as pertinent to the present investigation.

The Engineering Attitude Survey (EAS) is an instrument designed by Dr. William Larsen of the Metallurgy Department at Iowa State to provide information about engineering students' feelings toward their intended career. The EAS consists of 104 statements to be responded to in a five-point Likert format, varying from "strongly agree" to "strongly disagree". Since the psychometric characteristics of this instrument were not known, one of the secondary aims of this study was to validate the EAS. The responses were recorded on IBM answer sheets which were scored on the optical reader. This output was put into punched cards, retaining the five-point format.

The AVL, the ACL, and the SVIB are widely known instruments intended to measure personality and interest. A

factor analytic study of the Strong scales by Lewis, Wolins and Hogan (1965) found six clusters of scales which could be used to differentiate between graduation and dropout of engineering students. In general, however, as was indicated in the Introduction, work with such non-intellective predictors of academic performance has tended to produce disappointing results. It was theorized that one reason for this may be the nature of the scales which are generally used in interpreting the results from these instruments. While these scales have frequently been validated for certain applications, they seldom add anything in academic prediction. It was decided for this investigation to go beyond the usual scale scores and work as well with the items on each of these tests. For the ACL only the item scores were available, but for the AVL and the SVIB both the standardized scale scores and the item scores were included in the pool of independent variables. To get the item scores for these three tests, the answer sheets were transcribed into punched cards, retaining in all cases the original response format.

These procedures resulted in a total of 1197 "items" which were then used as independent variables in the present study. The source and number of these items is summarized in Table 3.

Table 3. Summary of independent variables

Test	Items	Scale scores	Factor scores	Total
MSAT	78	4	4	86
ENGL	90	3	4	97
MATH	65	1	4	70
ACT-B	46			46
EAS	104			104
AVL	120	6		126
ACL	300			300
SVIB	<u>399</u>	<u>59</u>	<u>—</u>	<u>458</u>
Total	1112	73	12	1197

Dependent variables

Following the spring quarter of 1969, by which time the subjects still in school had completed six quarters (two years) of college, complete transcripts of the academic career of each of the subjects in both groups were obtained from the Office of the Registrar. From these records the quarter GPA and the cumulative GPA were obtained for each subject for each of the quarters he had been enrolled at Iowa State. These data constitute the primary measures of academic performance for this study.

A criterion called "dropout" was defined by giving a score of "1" if he completed all six quarters and a score of

"0" if for any reason he failed to complete six quarters. There were 161 of the original 217 subjects who completed six quarters.

Additional indices which might be considered as measures of trends in performance were then generated. First, for each of the subjects who had a score of "1" on dropout, quarter GPA's were regressed against the six points in time assumed to be equally spaced. From this regression line the following quantities were computed: intercept (a), slope (b), slope squared (b^2), the product-moment correlation coefficient (r), estimate of error variance $\left[s_e^2 = s_y^2(1 - r^2) \right]$, and standard error of estimate (s_e). Second, it was decided to employ a modified curve-fitting procedure. Table 4 shows the coefficients of orthogonal polynomials which were used. An index of "fit" was defined as:

$$F_i = \sum (\text{GPA})_j \cdot C_{ij}$$

where $i = 1, 2, 3$ and $j = 1, 2, \dots, 6$. If the fit is good, the resulting sum (F_i) is large; if the fit is poor, the resulting sum is small (in absolute value). Only the linear, quadratic and cubic trends were investigated. The F_i provide a measure of which trend predominates, though no tests of significance were used to determine goodness of fit. A final index was coded "1" if the linear trend predominated, "2" if the quadratic, and "3" if the cubic.

Table 4. Coefficients of orthogonal polynomials for six treatments^a

	X_1	X_2	X_3	X_4	X_5	X_6
Linear (lin)	-5	-3	-1	1	3	5
Quadratic (quad)	5	-1	-4	-4	-1	5
Cubic (cub)	-5	7	4	-4	-7	5

^aAdapted from Hays (1965), Table VI, p. 682.

Table 5. Trend indicator intercorrelations*

	a	b	b^2	s_e	s_e^2	r	lin	quad	cub	max
a		63	04	-33	30	-59	-63	-23	-04	00
b			-17	-05	-05	94	100	08	04	05
b^2				06	05	-10	-17	-05	-21	-53
s_e					97	-01	-05	13	16	19
s_e^2						01	-05	16	19	16
r							94	06	-04	03
lin								09	-04	05
quad									16	04
cub										18

*Decimals omitted.

Table 5 shows the intercorrelations of these ten indices. As would be expected, b and lin correlate +1.00 with each other, since they are two ways of estimating the same thing. While this set of indices are obviously not all independent,

it was decided to retain nine of them for this investigation. Thus, the five orientation scores, the eleven GPA's, dropout, and the nine trend indices just described provided the 26 "criterion" measures used in this study.

Item-criterion correlation

The principal method of investigation was to correlate each of the "items" with each of the "criteria". The inter-correlations of the items with themselves were not of primary interest, since the total N was much too small to allow a factor analysis of this large pool of items.

A major difficulty was incomplete data; not all of the subjects had taken all of the tests, nor had they necessarily all taken the same group of tests. Data were complete for the three orientation tests. It would have been meaningless, given the nature of the other five tests, to have applied any kind of "missing data" technique. To solve this problem a special program was written in Fortran IV by the investigator. (This program will be found in Appendix B.) Its special feature is that it enters a subject's score into the covariance and correlation computation only if the subject has both the item score and the criterion score. With the use of this program the 30,922 correlations, with an average N of 186, required in this study were computed.

Since it was not the purpose of this phase of the investigation to distinguish between the experimental and

Table 6. Number of subjects having taken each test

Test	Number
MSAT	217
ENGL	217
MATH	214
ACT-B	174
EAS	181
AVL	163
ACL	166
SVIB	167

control groups, the data from both groups was pooled. Table 6 gives the number of subjects having the various measures. While this pooling sacrifices the possibility of cross-validation with the present sample, it was felt that cross-validation could come in future research.

Statistical Analysis

All computations required for the analyses of these data were made using standard or special Fortran programs run on the IBM 360/65 data processing system in the Computation Center at Iowa State University.

Comparisons of group differences in academic performance (Hypothesis I) were made using GPA's as criteria. Since the

N's in the two groups differed, analyses of variance were made using a regression technique previously delineated by the investigator (Klingensmith, 1968). Hypothesis II was tested by using a two-way contingency table in a chi-square analysis (see Hays, 1965, p. 590).

In the second and major phase of the investigation, no tests of significance as such for Hypotheses III and IV were made. Given the nature of the sample and number of correlations involved, simple tests of significance would have been meaningless. The dangers of applying textbook inferential statistics in educational research have recently been discussed by Coats (1970). The manner in which inferences can be drawn from the results of the present investigation will be discussed in the following section of this paper.

STUDY 1: RESULTS AND DISCUSSION

Since both phases of this investigation depend to some extent on the nature of the GPA distributions, these distributions for the six quarters were analyzed. Table 7 gives the frequency distributions of GPA's for the intervals shown. Using these data, the hypotheses that the GPA's were random samples from normal distributions were tested by computing the expected frequencies from a table of the normal distribution and applying chi-square goodness-of-fit tests to the results. For quarters two through five the hypothesis was rejected ($p < .01$). For the first quarter the observed distribution could have been a random sample from a normal distribution ($\chi^2 = 6.784$, $.05 < p < .10$); for the sixth quarter the same was true ($\chi^2 = 1.955$, $.70 < p < .80$). It was surprising that the sixth quarter distribution showed the closest fit to normality, since by that time any distortion due to restriction of range of ability should have been most extreme.

Thus we have evidence both for and against normality and we cannot conclude in general that these GPA's follow a normal distribution. However, it is well known that the F-test is robust with respect to violations of the assumption of normality, and we may safely use it for testing Hypothesis

Table 7. Distribution of grade point averages, all subjects

	Quarter					
	1	2	3	4	5	6
<u>GPA</u>						
4.00-3.50	7	14	12	9	13	13
3.50-3.00	32	24	30	35	23	22
3.00-2.50	60	51	45	33	39	45
2.50-2.00	62	60	59	42	51	46
2.00-1.50	40	36	32	33	31	29
1.50-1.00	13	23	13	12	11	13
1.00-0.0	1	4	11	7	8	1
Mean	2.46	2.38	2.38	2.42	2.42	2.49
St. dev.	.60	.71	.72	.77	.75	.69
N	217	212	202	171	176	169

I. Normality is required for product-moment correlation only if the correlation coefficients are to be subjected to tests of significance, which will not be the case in this study. The GPA's are seen by inspection to be essentially continuous and not severely skewed; we are therefore justified in using them in correlation analysis without applying a correction for attenuation.

Table 8. Means and standard deviations of five orientation scores

	Experimental (N = 115)	Control (N = 102)	Engineers (N = 1060)	University (N = 3200)
HSR	19.66 (13.06) ^a	20.47 (13.44)	22.01 (15.10)	23.58 (17.65)
ACT	26.44 (2.95)	26.42 (2.77)	26.29 (3.01)	25.27 (3.52)
MSAT	52.89 (11.67)	51.59 (11.90)	51.06 (11.95)	49.56 (13.00)
ENGL	53.57 (9.55)	50.91 (8.42)	51.41 (9.60)	52.56 (10.60)
MATH	41.67 (8.91)	40.66 (9.10)	41.28 (9.28)	39.49 (10.17)

^aStandard deviation.

Results: Comparisons of Experimental and Control Groups

Hypothesis I may be stated in the null form as follows:

There will be no difference in the mean GPA attained by engineering students who study design principles early in their college career and by those who follow traditional programs.

Table 8 gives the means and standard deviations of the scores attained on the five orientation tests. The scores for all engineers and all freshmen tested in the summer of

1967 are included in this table for comparison. The impression which these data immediately give is that on these five variables the two groups studied did not differ much from each other nor did either group differ much from the larger college population of which they were samples.

Tests for differences between the two groups in this study showed that in fact they did not differ other than by sampling variation on any of the five orientation variables. These tests actually came out of the regression analysis, though in effect they are t-tests for the difference between uncorrelated means.

Via the regression technique, analyses of variance were then made of the mean GPA's of the two groups for each of the six quarters. Table 9 summarizes the results of these analyses. In no quarter did the difference in mean GPA attain statistical significance at the .05 level. In the only case where it approached significance (third quarter), the mean of the control group was higher (2.48) than that of the experimental group (2.29). Since there were no differences in the two groups on the orientation tests, and since the covariances of the GPA's with these test scores were all positive, there was no reason to apply covariance analysis using the orientation scores as covariates.

Table 9. Summary of group comparisons

	Quarter					
	1	2	3	4	5	6
<u>Experimental</u>						
Mean	2.47	2.39	2.29	2.40	2.42	2.46
St. dev.	.57	.73	.78	.76	.75	.72
N	115	114	107	89	92	88
<u>Control</u>						
Mean	2.47	2.37	2.48	2.44	2.42	2.53
St. dev.	.67	.69	.67	.78	.75	.63
N	102	98	95	82	84	81
<u>F</u> -value	<1.0	<1.0	2.53	<1.0	<1.0	<1.0

Assuming that the GPA represents a valid index of the academic performance of a student, these results provide no evidence for rejecting the null hypothesis.

Hypothesis II may be stated in the null form as follows:
There will be no difference in the rate at which members of the two groups change curricula and/or drop out of college.

Table 10 summarizes the changes of curricula which occurred in both groups. Though 13 members of the experimental group and 11 members of the control group changed curricula more than once in the two years studied, only one

Table 10. Summary of curriculum changes

	Change	No-change	Totals
<u>Total changes of curriculum</u>			
Experimental	68 (59.35) ^a	47 (50.35)	115
Control	54 (52.65)	48 (44.65)	102
Totals	112	95	217
<u>Changes from engr. only</u>			
Experimental	48 (49.29)	67 (65.71)	115
Control	45 (43.71)	57 (58.29)	102
Totals	93	124	217

^aExpected values.

change (the first) is included in these data. The first part of the table considers all first changes without distinction. The quantities in parentheses are the expected frequencies computed from the marginal values (Hays, 1965, p. 590). The chi-square value computed from this two-way contingency table is 1.76 (1 df., .10 < p < .20). The second part of the table considers only those changes in which the student moved out of Engineering entirely. The chi-square value computed from this two-way contingency table is 0.13 (1 df., .70 < p < .80). Since neither of these values is

significant at the .05 level, there is no basis for rejecting the first part of the null hypothesis.

The information available is not sufficient to allow a precise statistical test of the second part of the hypothesis. Of the experimental group, 10 were formally dropped by the Engineering Academic Standards Committee, while only 4 of the control group were so dropped. An additional 13 of the experimental group and 14 of the control group left school voluntarily. It is not known whether these students continued their education at another institution, but it is evident in many cases that their withdrawal from Iowa State was not due to failing grades. There is, moreover, no apparent relationship between changing curricula and later withdrawal; about half of those who withdrew or were dropped had changed curricula and the other half had not. Thus there is no evidence for rejecting the second part of the null hypothesis.

Discussion

If one had hoped that the effects of the experimental program for training engineering students would be immediate and impressive, he would necessarily be disappointed by the above results. On the basis of the variables studied, there

is no apparent difference between the two groups, even after two years.

The originators of the experimental program felt that its two principal results would be increased student motivation for studying and a decreased attrition rate. Consequently, several cautions are in order in interpreting the above results. First, one may question whether the GPA is a measure sensitive enough to detect the increased academic competence which could be expected to follow from more intense study. Better grades could be expected to come from the innovative approach both indirectly, through increased motivation, and directly, as a result of better organization of engineering courses. Another source of increase, and one not envisioned by the originators of the program, might be expected from a combination of the Hawthorne and the experimenter effects: the subjects in the experimental group might do better because they knew they were expected to do better and because their teachers wanted them to do better.

At the same time, several factors militated against such increase. One of the aims of the experimental program was to break out of traditional grading molds and reward students for imaginative and creative thinking. The evidence suggests, however, that teachers tended to continue to grade "on the curve"; even if the experimental group became more proficient, there may have been imposed on them a grade

distribution which tended to obscure that increase. There is also the possibility that teachers who did not support the innovation tended to grade the subjects in the experimental program more stringently. It is problematic, then, just what combination of these factors is represented in the attained GPA's.

Second, there is the problem of how much increase in a measure like the GPA would be necessary in order to show meaningful results. It is here that classical inferential statistical procedures tend to lead us astray in educational research. With an N of approximately 200, a mean difference in GPA of only 0.27 would be significant at the .05 level. But this significant difference might not be "significant" at all. Some measure of strength of association between treatment and dependent measure must be computed and evaluated before we can say that the observed difference is meaningful. Consider the case of the third quarter, in which the difference in mean GPA approached statistical significance. The R^2 --an appropriate measure of strength of association here--related to this difference was .01, which means that only one percent of the variance was accounted for by the treatment effect. It is in this light that statistical significance by itself can lead to erroneous conclusions. On the other hand, failure to show significant difference does not necessarily mean that no differences exist.

The desire to reduce the attrition rate in the College of Engineering was the chief motivation which led to the establishing of the innovative program. From the point of view of utilizing manpower and resources, it is undesirable to have students begin a certain curriculum and then drop out before completing it. However, there is evidence that some beginning engineering students only think they want to become engineers, perhaps because of family traditions or other pressures, but without having any clear idea of what the engineering profession entails. For such a person, an innovative program which introduced the student early to a better understanding of the engineer's job might have the opposite effect than that intended: it might actually increase the number of dropouts. From the point of view of the student, this early discovery that he was in the wrong field would be to his advantage since he would not then waste time taking courses which might ultimately be of little use to him or obtain a degree in a field which he disliked. The results of this study show that, if anything, the attrition rate increased rather than decreased. More adequate pre-college prediction and guidance may be the only way around this difficulty.

The ultimate evaluation of the innovative program cannot yet be made. The present study is part of a continuing effort to evaluate this program. The final two years in the

college career of the subjects from both groups who remain at Iowa State is being subjected to similar analyses. A survey-type instrument, based on the results of the second part of the present study, is being designed to acquire additional information from these subjects. In the last analysis, of course, post-college criteria will be needed in order to evaluate the program completely.

STUDY 2: RESULTS AND DISCUSSION

It should be kept in mind in considering the results of this part of the project that the data for both the experimental and the control groups were pooled. Table 6 (see p. 36) gives the number of subjects who were administered each of the tests. Just as not all subjects took all tests, so not all criteria were available for all subjects. As a result, the N's of the item-criterion correlations vary from 125 to 217, with a median of about 165. For an N of 165, a correlation of 0.154 is significant at the .05 level; this provides a benchmark figure for reference.

Another point to note is that many of the items involved here are of the nature of artificially dichotomized variables: e.g., right-wrong on test items, yes-no on attitude items. Thus, since the GPA is essentially a continuous variable, we are often dealing with point-biserial correlations. These correlations were in fact computed using the product-moment formula. There is some disagreement among authorities as to whether the value of the point-biserial coefficient is numerically the same when computed from its own special formula as when computed from the product-moment formula. We follow Nunnally (1967, p. 120) in considering them to be interchangeable; in any case, the point-biserial

is an estimate of the product-moment r . The resulting numerical values are validity coefficients, and will be so treated here.

Results: Predictor-Criterion Relationships

A preliminary set of results concerned the factors which were extracted from the three orientation tests. These will be presented as a prelude to the item-criterion results.

Factor scales

MSAT Table 20 in Appendix C lists the MSAT items which were chosen for each factor on the basis of the criteria used (see p. 29). They are grouped under the names of the three logical subscales for ease of interpretation. Twelve of the MSAT items did not meet the criteria and are not included.

Table 11 shows the intercorrelations of the various MSAT scales and factors. It is obvious that Factor I is an Analogies factor, since it correlates .91 with the Analogies subscale. Factor IV seems to be a Reading factor, since it correlates .80 with the Reading subscale. Factors II and III are not so clearly related to the subscales, and an examination of their item content failed to reveal the unifying elements. The Same-Opposites subscale seems to relate least

Table 11. Intercorrelations of MSAT scales^a

	1	2	3	4	5	6	7	8
1. Total score		87	42	90	89	60	61	82
2. Reading			24	63	67	68	46	80
3. Same-Opposites				41	35	27	18	41
4. Analogies					91	45	65	60
5. Factor I						50	41	59
6. Factor II							24	37
7. Factor III								40
8. Factor IV								

^aDecimals omitted.

to any of the other scales. The intercorrelations of the four factor scores show that they are less independent than might have been expected, given the stringency of the criteria for deriving them. It should be noted that the Reading and Analogies subscales, and Factor I and Factor IV scales, all correlate highly with total score. Reliability estimates of the MSAT, using the KR-20 estimate, are usually in the range of .90 and up, indicating a very homogeneous and internally consistent test. The above results tend to confirm this, and indicate that the test is getting at the ability to recognize the meaning of words and to think analogically with them.

ENGL Table 21 in Appendix C lists the ENGL items which comprise the four extracted factors. They are grouped

Table 12. Intercorrelations of ENGL scales^a

	1	2	3	4	5	6	7
1. Total score		72	94	75	69	56	62
2. Effectiveness			45	41	89	26	26
3. Mechanics				78	47	60	67
4. Factor I					35	21	32
5. Factor II						23	25
6. Factor III							24
7. Factor IV							

^aDecimals omitted.

under the names of the two logical subscales for ease of reference. Fifteen of the ENGL items did not meet the criteria and are not included.

Table 12 shows the intercorrelations of the various ENGL scales and factors. Factor I, which correlates .78 with the Mechanics subscale, seems to be a Mechanics factor. Factor II, which correlates .89 with the Effectiveness subscale, is evidently an Effectiveness factor. While Factors III and IV are, like Factor I, composed almost entirely of items from the Mechanics part of the test, neither of them correlates as highly with the Mechanics subscale as does Factor I. Regarding their correlations with the subscales and with total score, Factors III and IV appear to be very similar; yet they correlate only .24 with each other. This suggests an interesting overlap of variances which might profitably be

Table 13. Intercorrelations of MATH scales^a

	1	2	3	4	5
1. Total score		62	77	56	61
2. Factor I			26	26	27
3. Factor II				33	34
4. Factor III					20
5. Factor IV					

^aDecimals omitted.

investigated further. The intercorrelations of the four factor scores show that they are relatively independent. The correlation of .94 between total score and the Mechanics subscale may be explained by the fact that this subscale constitutes two-thirds of the test.

MATH Table 22 in Appendix C lists the item content of the Math factor scales. Table 13 shows the intercorrelations of the MATH factor scales with total score. The highest correlation with total score is that of Factor II. Since the items in this scale are all from the latter part of the test, this may indicate a Completion factor: students who complete more items tend to get higher scores, even though this test is scored using a formula to correct for guessing (rights minus one-quarter wrongs). This interpretation is supported by the fact that this scale tends to correlate more highly than the others with such measures of academic proficiency as

high school rank. There were 18 items in this test which did not meet the criteria for inclusion in a factor scale. The intercorrelations of the factors show that they are relatively independent. Inspection of the test items indicates that Factor I is an Elementary Operations and Basic Algebra factor. Similarly, Factor III is a Word Problems factor involving the ability to ferret out the required manipulation from a statement. Factor IV is not so clear, but it seems to be a Geometry factor requiring the ability to visualize relationships of angles and lengths.

Scale-criterion relationships

It will be recalled that among the various "items" in this study, certain ones are actually the total and subscale scores just described. Since these are the aspects of the given instruments generally used in research, their relationships to the selected criteria will be considered next. It may be noted that the program used to compute these correlations displayed only those which were larger in absolute value than 0.10 (see Appendix B). The six quarter GPA's, the dropout criterion, and the three trend indicators were found to be the most interesting criteria.

Academic scales Table 14 shows the correlations between the various scales on the three orientation tests and the selected criteria. As would be expected, the total score

Table 14. Academic scales: validity coefficients

	Q1	Q2	Q3	Q4	Q5	Q6	D0	Lin	Quad	Cub
<u>MSAT</u>										
Total	42	30	22	21	12	13	--	-12	--	--
Reading	37	27	18	18	--	13	--	-14	--	--
Same-Opposites	11	--	--	--	--	--	--	--	--	--
Analogies	36	27	25	20	14	10	--	-10	--	--
Factor I	37	31	21	21	12	--	--	-16	--	--
Factor II	30	16	13	10	--	--	--	-14	--	--
Factor III	28	21	17	11	--	10	--	--	--	--
Factor IV	30	19	12	15	--	17	--	--	--	--
<u>ENGL</u>										
Total	36	26	11	19	15	18	--	--	--	--
Effectiveness	26	24	16	13	--	11	10	--	--	--
Mechanics	34	22	--	19	16	19	--	--	--	--
Factor I	27	14	--	11	--	16	--	--	--	--
Factor II	26	23	17	21	11	17	--	--	--	--
Factor III	15	15	--	--	--	--	12	--	--	--
Factor IV	18	--	--	--	15	16	-10	--	--	--
<u>MATH</u>										
Total	32	22	14	14	13	11	--	-14	--	--
Factor I	18	14	--	12	14	--	--	--	--	-10
Factor II	26	16	11	--	--	--	--	--	14	--
Factor III	17	16	15	22	18	--	--	--	-16	--
Factor IV	20	19	11	--	--	13	--	-11	10	--

on each of these tests is the best predictor of first quarter GPA. One or more of the subscales in each test, however, show validity coefficients of the same order of magnitude as the total score. The validity of the total score decreases regularly across quarters, as has been found in other studies.

Some of the sub- or factor scales show less attenuation over time. For example, the ENGL Factor II and the MATH Factor III seem to predict about equally well across five quarters. Only the ENGL scales have any relation to the dropout criterion (and this relation is very tenuous), yet they fail completely to predict trend. Only the MATH scales predict quadratic trend, and the MSAT and MATH predictions of linear trend are moderate. None of these scales seems to be related to the cubic trend.

Non-academic scales Since scale scores were available for only the ACL and SVIB non-academic measures, only these are displayed in Table 15.

There are only six scales on the AVL, and all of them are shown even though some of them have little relation to the criteria. The Theoretical scale might be expected to predict GPA, but at least for this group of engineers it does not. The Religious scale is the most consistent predictor of GPA, though the correlations are moderate. It also predicts two of the three trend indices. It is interesting that the Political scale predicts all three of the trend indices about equally well. None of the six scales predicts dropout.

While all 59 of the SVIB scales showed some relationship to at least one of the selected criteria, only those in Table 15 seemed to involve any consistent relationship with

Table 15. Non-academic scales: validity coefficients

	Q1	Q2	Q3	Q4	Q5	Q6	D0	Lin	Quad	Cub
<u>AVL</u>										
Theoretical	--	--	--	--	-11	--	--	--	--	--
Economic	--	--	-16	--	--	--	--	--	--	--
Aesthetic	--	--	--	--	--	--	--	--	--	--
Social	-12	--	--	--	--	-11	--	-11	--	--
Political	--	--	--	--	--	12	--	14	13	17
Religious	11	17	13	14	12	--	--	--	-14	-11
<u>SVIB</u>										
Veterinarian	-19	-20	-13	--	--	--	--	--	--	-16
Psychologist	19	21	19	--	--	--	--	-13	-12	--
Production manager	-14	-13	-14	--	--	--	--	18	--	-16
Carpenter	-11	-11	-13	--	--	--	--	--	--	-12
Policeman	-23	-23	-20	--	--	--	--	18	11	--
Librarian	16	21	15	--	--	11	--	-13	--	23
Pharmacist	-20	-19	-16	--	--	23	--	--	-11	-21
Mortician	-32	-31	-17	-16	--	-11	-15	13	12	--
Real estate salesman	-27	-32	-15	-18	-13	--	-21	--	12	--
Life ins. salesman	-21	-23	-14	-19	--	-16	-21	--	--	--
Academic Ach.	27	28	14	--	--	--	10	-19	-15	12
Engineer	05	02	-04	13	06	-04	02	09	-14	09

several of the criteria. As a point of interest, the exact correlations of the Engineer scale with all ten criteria are also shown. The Strong scales are scored in such a way that a high score indicates similarity with a norm group. Consequently, the negative correlations which predominate in

Table 15 are to be interpreted to mean that similarity to the profession named by the scale is inversely related to GPA. Most of the scales shown seem to predict GPA quite consistently over the first three quarters. Many of the scales in this group have validities equal to those of the academic scales. Also, many of the scales listed predict the various trend indicators better than did any of the academic scales. The Academic Achievement scale is often interpreted as a measure of persistence, an interpretation which does not seem to be verified here, at least beyond the second quarter. The Engineer scale, which might seem to have special relevance to this group of subjects, is related only to fourth quarter GPA and to quadratic trend.

Item-criterion relationships

Here we are considering items in the strict sense, i.e., the individual items on the various tests administered. The amount of data here is very large and does not admit of easy summarization. Some of the more interesting results will be considered, what is "interesting" being of course a matter of judgment.

Academic test items Table 16 shows the correlations between the selected criteria and certain items on the three academic orientation tests. The criterion for inclusion in this table was that the item seemed to involve some

Table 16. Academic test items: validity coefficients

	Q1	Q2	Q3	Q4	Q5	Q6	DO	Lin	Quad	Cub
<u>MSAT</u>										
38	23	16	21	16	--	16	--	--	--	--
45	17	15	17	17	15	24	--	--	--	15
46	22	18	12	13	11	15	--	--	--	--
74	25	21	18	16	--	10	--	-10	--	--
<u>ENGL</u>										
15	18	18	20	--	14	11	12	--	--	--
49	19	20	10	16	--	20	--	--	--	--
69	14	14	12	18	14	14	--	--	--	--
72	19	16	16	--	18	20	--	--	--	--
88	23	11	21	13	--	14	--	-11	--	--
<u>MATH</u>										
29	26	20	16	21	14	--	--	--	-18	-12
37	26	22	16	--	10	10	--	--	--	--
55	19	17	11	--	11	13	--	--	--	--

consistent relationship with several of the criteria. The pattern displayed in this table is quite similar to that in Table 14 (see p. 55), and the validity coefficients of the single items are of the same order of magnitude as are those of the scales. This suggests that, with the exception of the first quarter GPA, as much information may be obtained with a few items as is obtained from the whole test. These items, however, do little in the way of predicting dropout or performance trend.

Non-academic test items Table 17 shows the correlations between the selected criteria and certain items on the five non-academic measures. The criterion for inclusion here was the same as that for the previous table. The interesting thing about these results is that the validity coefficients are as large as, and in many cases larger than, either the scale score validities or the academic test item validities. This again is somewhat surprising in view of previous research which has tended to find no relationship between non-intellective predictors and measures of scholastic achievement.

Table 18 shows which items from each of the eight tests proved to be the best single predictor of each of the selected ten criteria. The nature of the items in these three tables will be discussed in the next section of this report.

Conclusion

Since Hypotheses III and IV are not being subjected to statistical verification, it is not necessary to state them in the null form.

Hypothesis III postulated the existence and discernibility of patterns of academic performance. Of the various trend indices investigated in this study, the linear, quadratic, and cubic trend sums seem to be the most interpretable ones. The results displayed in Tables 14 through 17 indicate

Table 17. Non-academic test items: validity coefficients

Item no.	Q1	Q2	Q3	Q4	Q5	Q6	D0	Lin	Quad	Cub
<u>ACT-B</u>										
3	18	25	17	12	--	15	10	-11	--	17
22	17	16	16	10	14	--	--	--	--	--
28	15	17	23	18	26	11	--	12	--	--
<u>EAS</u>										
61	--	26	17	23	17	26	17	--	--	16
76	17	14	14	--	-10	-16	--	-11	--	--
<u>AVL</u>										
1	11	12	10	--	18	15	--	--	--	--
60	-13	-22	-19	-15	-17	--	-14	--	--	--
72	-11	-16	-12	--	-11	-12	--	--	--	-15
<u>ACL</u>										
54	19	14	21	11	--	11	16	--	--	12
129	12	21	14	16	17	16	16	--	--	13
141	-24	-30	-20	-19	-22	--	-16	--	--	--
173	15	14	13	--	16	12	22	--	--	--
180	-14	-15	-17	-17	--	18	--	--	--	-14
185	15	16	13	11	16	14	--	--	--	--
244	-13	-15	-20	-17	-12	--	--	--	12	--
261	-19	-17	-18	-18	-15	-19	--	--	13	--
<u>SVIB</u>										
7	12	10	16	12	12	13	--	--	--	--
145	25	21	16	12	13	--	--	-15	--	--
168	28	18	34	13	10	14	--	--	-12	--
195	-12	-15	-11	-18	-22	-11	-15	-15	12	--
206	--	11	20	13	11	17	--	--	--	--
235	-20	-11	-21	-12	--	-11	-17	--	--	--
270	-20	-17	-18	-10	--	-22	-16	--	--	--
287	21	30	21	20	11	18	--	--	-12	14
316	-17	-26	-15	-15	-11	--	--	--	--	--
317	18	31	26	19	11	--	13	--	-23	11
342	20	17	20	13	11	--	14	--	-12	--
345	25	24	23	16	14	15	15	--	-10	--
360	16	12	23	13	13	--	18	--	--	--
365	15	14	15	15	15	15	13	--	--	--
367	-12	-15	-13	-16	-20	-18	--	--	--	--
372	-43	-29	-22	-25	-34	-20	--	--	--	26
373	-23	-17	-20	-19	-12	-13	--	--	--	--

that it is at least as feasible to predict performance trend as it is to predict GPA. To this degree, Hypothesis III may be said to be moderately supported by the present research. It is entirely possible that other and better measures of trend will be developed.

Hypothesis IV postulated a relationship between performance trends and pre-college measures. To demonstrate the method by which this hypothesis might be verified, the following procedure was used. The 217 subjects in the study divided themselves into four groups of about equal size, on the basis of linear, quadratic, and cubic trend, and dropout. Frequency counts were made for each of these groups on responses to the ACL. If a group chose or omitted a response with a ten percent higher frequency than did any other group, that response was used to describe that group. The adjectives so assigned are shown in Table 19. The upper part of each list gives the items predominantly chosen; the lower part gives the items predominantly omitted. From this listing it can be seen that the self-descriptions of the four groups do differ. The linear group, which performed in a regular fashion, describes itself as intelligent, self-controlled, stable and steady. The quadratic group describes itself as disorderly, nervous and suggestible. The cubic group describes itself as persevering--perhaps in spite of its ups and downs; but it is best defined by its reluctance to

Table 19. ACL items descriptive of the four performance groups

Linear	Quadratic	Cubic	Dropout
14 argumentative	66 disorderly	151 mild	41 confident
87 fickle	77 emotional	173 persevering	76 egotistical
119 immature	103 good-looking		78 energetic
124 individualistic	146 mannerly		79 enterprising
125 industrious	158 nervous		88 flirtatious
127 informal	168 outspoken		90 forceful
132 intelligent	178 pleasure-seeking		92 forgetful
155 moody	228 show-off		105 handsome
160 obliging	254 suggestible		108 hasty
164 optimistic			117 idealistic
171 peaceable			120 impatient
177 pleasant			137 irresponsible
203 reserved			150 methodical
214 self-controlled			152 mischievous
220 sensitive			159 noisy
222 serious			175 pessimistic
240 soft-hearted		84 fault-finding	189 queer
246 stable		89 foolish	197 rebellious
247 steady		123 indifferent	198 reckless
259 tactful		130 initiative	211 sarcastic
266 thoughtful		134 interests wide	229 shrewd
278 understanding		162 opinionated	263 tense
		163 opportunistic	274 unambitious
		179 poised	282 uninhibited
		193 rational	293 weak
		219 selfish	
	116 hurried	239 sociable	
169 painstaking	188 quarrelsome	244 spontaneous	83 fair-minded
183 precise	191 quiet	258 sympathetic	170 patient
190 quick	231 silent	262 temperamental	195 realistic
216 self-pitying	252 stubborn	291 warm	217 self-punishing

describe itself. The dropout group seems to display many contradictory facets, e.g., confident yet pessimistic, forgetful yet shrewd.

Future extensions of the present study will attempt to relate other clusters of non-intellective variables to these performance trends. For the present study, these results offer some support for Hypothesis IV: performance trends as here defined are related to both intellective and non-intellective pre-college variables (see Table 18).

Discussion

Even though Study 2 was cast in the form of two research hypotheses to be verified, it had more the purpose of hypothesis generation than hypothesis testing. The subjects in the study were beginning engineering students in a single College of Engineering; it is not meaningful to try to generalize the results to other populations. But the techniques which have been developed and utilized here can and should be applied to the study of other populations.

Non-intellective predictors

Perhaps the most important result of this study is the demonstration of the feasibility of relating non-intellective predictors to criteria of academic performance. This was

Table 18. Test items which correlate maximally with criteria

	MSAT	ENGL	MATH	ACT-B	EAS	AVL	ACL	SVIB
Q1	57 ^a (31) ^b	89(23)	43(33)	8(-23)	30(25)	113(-19)	55(-27)	168(28)
Q2	34(28)	49(20)	36(25)	3(25)	61(26)	105(26)	141(-30)	317(31)
Q3	43(22)	88(21)	58(16)	28(23)	39(-22)	59(20)	113(25)	168(34)
Q4	76(28)	13(19)	9(23)	28(18)	4(29)	87(20)	242(-23)	34(-28)
Q5	45(15)	11(-19)	27(18)	28(26)	90(-33)	49(21)	138(26)	372(-34)
Q6	56(-23)	78(24)	33(20)	14(-28)	80(30)	31(22)	10(23)	183(28)
DO	34(17)	8(17)	47(15)	10(-17)	90(-24)	31(22)	274(-24)	250(-25)
Lin	49(-24)	54(-21)	43(-21)	14(-27)	56(-27)	31(26)	254(29)	183(30)
Quad	43(-15)	52(-22)	5(23)	29(17)	60(-23)	71(32)	242(28)	166(28)
Cub	76(-20)	19(-15)	3(-18)	31(-19)	13(-24)	83(-24)	296(-26)	372(26)

^aItem number.

^bCorrelations; decimals omitted.

accomplished by looking not only at the established scale scores on some typical personality, biographic, and interest measures, but also by studying the individual items on these measures. Most such measures were developed and validated for specific purposes, which they often serve quite well. But the use of these measures in a global fashion as predictors of academic achievement has usually produced less than satisfactory results. The low internal consistency of such measures indicates that they are not unifactor tests. Factors derived from these tests have sometimes proved useful for prediction. The present research goes a step further in looking at the individual items and in relating them to academic criteria. In this process, some interesting results appear.

If we examine the items on the non-intellective measures used in this study which correlate best with the selected criteria, the following picture of the successful Iowa State engineering student appears. From the ACT-B we find that he is one who in high school has had notable achievements in science outside the classroom and not as part of course assignments (item 22); he has come to Iowa State because he is concerned with quality of faculty (item 28) and with a desirable intellectual atmosphere (item 31); in college, he tends to be studious and to avoid certain types of extra-curricular activities (item 14).

The pertinent items on the EAS help to fill out this picture. The successful engineering student has a dogged devotion to his job. He is not concerned that his profession be a thrilling experience or that it bring him personal satisfaction (item 61); in fact, he is willing to sacrifice personal satisfaction in order to succeed (item 13). He thinks he has little to learn from history (item 39), and he tends to be anti-union (item 60). He is independent and prefers to follow his own views rather than those imposed on him (item 56); however, he does not expect to have much control over the use of his products as an engineer (item 80). He is a realistic, if somewhat pessimistic or fatalistic, individual.

Despite the fact that the AVL has been criticized for confusing values with interests, it adds another dimension to the picture of this student. The Religious scale is the best predictor of academic performance here, and this scale is generally found to identify the person who needs to come to terms with the cosmos, to find unity in life's experiences. Our student considers the discovery of truth more important than its application (item 1); he considers it more important to train his children in religion than in athletics (item 60); and he is opposed to an overemphasis on the practical (item 72). He is a practical man but not at the expense of a larger view.

On the SVIB, engineering students tend to score high on the physical science scales. This was true of these subjects, yet these are not the scales which are related to academic success in this study. For example, the Engineer scale is a poor predictor for this group. The successful student here places a high value on a college education; the negative scale correlations are distributed among those occupations for which the norms were not college graduates. On the positive side he resembles the sole proprietor, the small business man, though he would never say he wants to be a small business man. He seems to have enrolled in engineering because this satisfies a need to make a life decision, to adopt a satisfactory life style; if he changes curricula, it is to take up some form of business administration, which evidently provides a similar instrumentality for the satisfaction of a need. His choices must be very clearcut; he wants to be an engineer, but he has few romantic illusions about what that will entail.² He would dislike being an athletic director (item 7); he is not fond of visiting night clubs (item 168) or meeting people (item 206). He likes to argue (item 195), and he prefers the company of conservative and orderly

²The writer is indebted to Professor Donald Zytowski of the Student Counseling Service at Iowa State University for an interesting discussion of the Strong scales. The views expressed here, however, are not necessarily in agreement with his.

people (items 235 and 270). He would be much opposed to having to sell the product he designs (item 287). There is little in this picture to contradict the evidence from the other measures. It is perhaps noteworthy that on none of these measures does the successful student show any interest in aesthetic values; or, if he does, those interests have no relation to his success.

Trend indicators

The results of this study regarding trend-of-performance indicators were not as satisfactory as had been hoped. Of the nine indicators defined and investigated, only the linear, quadratic and cubic sums were found to have relationships to pre-college variables. But the possibilities were not exhausted. The relationships of these three indicators, as well as of the dropout criterion, with self-descriptions on the ACL were interesting findings and tend to support the theory that patterns of performance exist and can be related to other variables.³ It is expected that relating these criteria to clusters of items on other measures will further specify the nature of these indicators.

³The ACL includes such adjectives as arrogant, coarse, cowardly, effeminate, foolish, irresponsible, and spineless. It is curious that there was only one item in the 300, however, which no subject applied to himself--infantile.

There remains more information to be extracted from the data. For example, only those predictors which showed some consistency across time were considered in this study; there are many cases of predictors which correlate positively with a criterion over several quarters and then correlate negatively with the same criterion over subsequent quarters. In other cases, a predictor is seen to relate to a criterion for awhile and then ceases to relate, and vice versa. From such phenomena it may be possible to identify developmental changes in performance. These phenomena fit nicely into the general developmental model proposed by Schaie. While the present research does not prove the existence of person-environment interactions in the college experience, such interactions appear more clearly as the most likely explanation of the relationships between personality traits and academic performance.

Significance

The fact that the two groups were seen in the first study to be quite similar led to the necessity of making a choice for the second study. One choice would have been to divide the subjects into two equal-sized groups, perform the second study on one of these, and then validate the results on the other group. Or a double cross-validation design would have been used. Such a choice would have implied the

existence of specific hypotheses to be tested. Since the principal concern was rather to generate hypotheses, it was decided to use all of the subjects for the one investigation and leave validation to future research. It would be meaningless to combine the selected items revealed by this study in multiple regressions on the criteria since the results would necessarily be positive. Rather, the present results are being used to design an instrument which will be validated on other groups of engineers.

A word should be said about interpreting the present results statistically. As stated earlier, for an N of 165 a correlation of about 0.15 is significant at the .05 level. This must be understood to mean the following: with repeated random samples of size 165 from a population in which the population correlation coefficient, rho, is equal to zero, a sample correlation of .15 or greater can be expected once in twenty samples. Since this study involved over 30,000 correlations, about 1500 of them could be significant by chance, and there is no way to know which 1500 they are. For this reason the significance of the obtained correlations--which, in fact, are in the range of 0.15--is not discussed. It might be noted, however, that if a certain predictor correlates across time with a criterion, this fact suggests a kind of quasi-validation of the correlation.

Implications

The results of this study suggest guidelines in two areas: design of future research, and guidance and counseling.

If academic performance is related to non-intellective as well as intellective variables, it will be necessary to design person-environment studies which take into account aptitude-trait-interest-environment interactions. Aptitude alone obviously is not an adequate predictor of academic accomplishment. Since personality traits and individual interests are shown to be related to academic performance, they must also be factored into the equation. If the successful engineering student is as he has been described above, then the environment or "treatment" he was subjected to in the college experience must have in fact been compatible with his needs as this study has delineated them. In order to bring out the best in a student, a treatment must be concerned not only with his intellectual aptitude but also with his aspirations and his goals.

When and if the findings of the present study are validated, they will provide the guidance counselor with improved tools for determining the probable success of a student in a given field. A single instrument based on the findings of this and similar studies should be able to give the counselor better information than he currently obtains from a battery of personality and interest inventories, at least as regards

probable academic performance. There is little hope of understanding completely the full complexity of human behavior; there is hope of finding more parsimonious ways of describing and predicting that behavior.

SUMMARY

This study had a dual purpose. The first purpose was to evaluate the effectiveness of an innovative engineering curriculum in reducing dropout rate and in motivating the students to perform better academically. An experimental group of 115 entering freshman engineering students participated in the new program; a matched control group of 102 took the regular program. The dependent variables were GPA's for the first six quarters of college experience and counts of the number of students who changed curricula or dropped out of school. When these data were subjected to one-way analysis of variance and chi-square analysis respectively, no differences were found between the two groups. The research hypotheses that the innovative program would improve interest and reduce attrition were not verified. An explanation of why the new program failed to produce results was given, and suggestions for further evaluating the long-range effects of the program were made.

The second purpose was to seek new approaches to the problem of predicting academic performance in college. The subjects in both the above groups had been given a series of aptitude, personality, and interest tests at the beginning of their freshman year. These test results, plus scales and items derived from them, were used as independent variables.

Since the results of the first study showed the two groups to be similar, the data for all subjects was pooled for the second study. The large pool of correlations between independent and dependent variables was examined for the information it might provide about the nature of academic performance in college. It was shown that certain non-academic measures are able to predict academic achievement as well as ability measures do. It was further shown that performance trends are identifiable and can be related to pre-college variables. The implications of these findings for freshman testing programs and for counseling and guidance work were discussed. Plans for further research along these lines were described.

REFERENCES

- Astin, A. W. An empirical characterization of higher educational institutions. Journal of Educational Psychology, 1962, 53, 224-235.
- Astin, A. W. Comment on a student's dilemma: big fish-little pond or little fish-big pond. Journal of Counseling Psychology, 1969, 16, 20-22.
- Astin, A. W., & Holland, J. The environmental assessment technique: a way to measure college environments. Journal of Educational Psychology, 1961, 52, 308-316.
- Bracht, Glenn H. The relationship of treatment tasks, personological variables and dependent variables to aptitude-treatment interactions. Ph.D. Thesis. University of Colorado, Boulder, 1969.
- Brown, Fred G., & Scott, D. The unpredictability of predictability. Journal of Educational Measurement, 1966, 3, 297-301.
- Burket, G. R. A study of reduced rank models for multiple prediction. Psychometric Monographs, 1964, No. 12.
- Chickering, A. W., McDowell, J., & Campagna, D. Institutional differences and student development. Journal of Educational Psychology, 1969, 60, 315-326.
- Coats, Wm. Significant differences: A case against the normal use of inferential statistical models in educational research. Educational Researcher, 1970, 21, 6-7.
- Cronbach, L. J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.
- Cronbach, L. J., & Snow, R. E. Individual differences in learning ability as a function of instructional variables. Stanford University, March, 1969.
- Fishman, J. A. & Pasanella, A. K. College admission-selection studies. Review of Educational Research, 1960, 30, 298-310.

- Fleishman, E. A., & Ellison, G. D. Prediction of transfer and other learning phenomena from ability and personality measures. Journal of Educational Psychology, 1969, 60, 300-314.
- Garrett, Harley F. A review and interpretation of investigations of factors related to scholastic success. Journal of Experimental Education, 1949, 18, 91-138.
- Hays, Wm. L. Statistics for Psychologists. New York: Holt, Rinehart & Winston, 1965.
- Horst, Paul. Differential prediction in college admissions. College Board Review, 1957, 33, 19-23.
- Humphreys, L. G. The fleeting nature of the prediction of college and academic success. Journal of Educational Psychology, 1968, 59, 375-380.
- Jensen, A. R. Social class, race and genetics: implications for education. American Educational Research Journal, 1968, 5, 1-42.
- Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.
- Juola, A. E. Freshmen level ability tests and long range prediction. National Council on Measurement in Education paper, February, 1963.
- Juola, A. E. Prediction of successive terms performance in college from tests and grades. American Educational Research Journal, 1966, 3, 191-197.
- Juola, A. E. Illustrative problems in college-level grading. Personality and Guidance Journal, 1968, 47, 29-33.
- Kessen, W. Research design in the study of developmental problems. In P. Mussen (Ed.), Handbook of research methods in child development, New York, Wiley, 1960.
- Klein, S. P., Rock, D. A., & Evans, F. E. The use of multiple moderators in academic prediction. Journal of Educational Measurements, 1968, 5, 151-160.
- Klingensmith, J. E. Computational procedures for the regression analysis of variance. Mimeo. paper. Iowa State University. 1968.

- Lavin, D. E. The prediction of academic performance: a theoretical analysis and review of research. New York: Russell Sage Foundation, 1965.
- Lewis, E. C., Leroy Wolins, and John Hogan. Interest and ability correlates of graduation and attrition in a college of engineering. American Educational Research Journal, 1965, 2, 63-74.
- Lunneborg, C. E. Biographic variables in differential vs. absolute prediction. Journal of Educational Measurement, 1968, 5, 207-210. (a)
- Lunneborg, C. E. Use of factor scores in differential prediction of academic success. Journal of Educational Measurement, 1968, 5, 297-300. (b)
- Lunneborg, P. W., & Lunneborg, C. E. The differential prediction of college grades from biographic information. Educational and Psychological Measurement, 1966, 26, 917-925.
- Mayhew, L. B. Non-test predictors of academic achievement. Educational and Psychological Measurement, 1965, 25, 39-46.
- Meehl, Paul E. Clinical vs. Actuarial Prediction. Minneapolis: University of Minnesota Press, 1954.
- Menne, J. W., & Klingensmith, J. E. Grade distributions in multi-section courses. Mimeographed paper. Iowa State University, Ames, Iowa, 1968.
- Mischel, Walter. Continuity and change in personality. American Psychologist, 1969, 24, 1012-1018.
- Mitchell, J. V. Education's challenge to psychology: the prediction of behavior from person-environment interaction. Review of Educational Research, 1969, 39, 695-721.
- Morgan, Paul. A freshman program for engineering students: Committee Report. Mimeographed paper. Iowa State University, 1966.
- Nunnally, Jum C. Psychometric Theory. New York: McGraw-Hill, 1967.

- Ohnmacht, Fred W. Correlates of change in academic achievement. Journal of Educational Measurement, 1968, 5, 41-44.
- Pace, C. R. and Stern, G. C. An approach to the measurement of psychological characteristics of college environments. Journal of Educational Psychology, 1958, 49, 269-277.
- Sanford, Nevitt (Ed.) The American College. New York: Wiley, 1962.
- Schaie, K. W. A general model for the study of developmental problems. Psychological Bulletin, 1965, 64, 92-107.
- Stern, G. C. Characteristics of the intellectual climate in college environments. Harvard Educational Review, 1963, 33, 5-41.
- Werts, C. E. and Watley, D. J. A student's dilemma: big-fish-little pond or little fish-big pond. Journal of Counseling Psychology, 1969, 16, 14-19.

ACKNOWLEDGMENTS

This project could not have been carried out without the help of many persons. I wish to express my sincere thanks to Professor Edwin C. Lewis for his help and encouragement throughout my graduate program, especially for his patience when I was moving very slowly and his forbearance when I was moving very fast;

To Professors Wilbur Layton, Virgil Lagomarcino, Fred Brown and James Walsh, for serving on my graduate committee and for their many helpful suggestions;

To Professor John Menne, without whom none of this would have happened;

To Professor William Larsen of the College of Engineering and to the Registrar at Iowa State University, for permission to use some of the data on which this project depended;

To Mrs. Amy Ladman, for the countless hours she spent in the painstaking checking of data, and to all the girls in the Counseling Service for their help;

To many friends and well-wishers, for the confidence and encouragement which saw me through those times when the prize did not seem worth the rigors of the race.

APPENDIX A: FRESHMAN PROGRAM FOR ENGINEERING STUDENTS

While our modern technological age has great need for imaginative and perceptive engineers, many schools of engineering find themselves turning out only well trained technicians. The traditional engineering course lays heavy stress on foundational courses in science and mathematics; it lays little stress on synthesis and design until the very end of the course of study, and it often fails to acquaint the student with the nature of the profession and the role of the engineer as a professional man.

On November 23, 1964, Dean George R. Town appointed a committee to make a review of the basic engineering program at Iowa State. Four months later this first committee made its report to Dean Town. There were wide differences of opinion among the committee members about what an "ideal" freshman program should contain. The report contained a number of proposals for the implementation of a revised program.

After reviewing this report, Dean Town appointed Prof. Paul Morgan head of a second committee charged with formulating specific plans to implement some of the recommendations of the first committee. The work of this second committee resulted in two important recommendations: first, that the existing courses be thoroughly revised and

"streamlined"; and second, that a new sequence of courses be added at the freshman level which would add a large body of new material to the engineering program. To quote from the committee report: "It is not only hoped that they (the new courses) will motivate a freshman student but it is expected that they will inform him about engineering in a rather profound and sophisticated way and that they will encourage the qualities of curiosity, originality, and independence which will vitalize his entire professional career." (Morgan, 1966)

The experimental courses are given below:

Fall Quarter

- | | |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Engr. 190V | Introduction to Engr.

To acquaint the student with engineering and to develop an understanding of the relationship between engineering and the whole of life and knowledge. (1 credit) |
| Engr. 190W | Freshman Engr. No. 1

To provide the engineering student with the basic tools necessary to solve engineering problems and to present them analytically and graphically. (3 credits) |

Winter Quarter

- | | |
|------------|-----------------------------------------------------------------------------------------------|
| Engr. 190X | Freshman Engr. No. 2

To develop the thought process of engineering design. (3 credits) |
|------------|-----------------------------------------------------------------------------------------------|

Spring Quarter

Engr. 190Y

Freshman Engr. No. 3

To apply the thought process of design
(materials, systems, structures, etc)
(3 credits)

The committee then stated that this program should be considered as an educational experiment to see if its goals could be met, and it recommended that a pilot section of about 120 freshman students should be established.

This program and the committee's recommendations were approved by Dean Town, and it was decided to initiate the experimental section in the fall of 1967. Dr. William Larsen of the Metallurgy Department, a member of the second committee, developed a special instrument designed to measure the attitudes toward engineering of the students who would be in the experimental program. Dr. Fred Brown, of the Student Counseling Service, was brought in as special consultant to design the evaluation phase of the experimental program. It was his suggestion that a control group of students taking the regular program be added in order to allow comparisons of the results of the experimental program.

APPENDIX B: FORTRAN PROGRAM FOR COMPUTING ITEM-CRITERION CORRELATIONS

On the following pages is shown the computer printout of the special program written by the investigator to compute the numerous correlations required in the second phase of this study.

This program has three features. First, it utilizes integer arithmetic in order to reduce rounding error in the computations. All sums and sums of squares are accumulated in integer and converted to floating point only for the final divisions and square roots.

Second, it incorporates a multiple-access feature which allows the data to be re-read as often as necessary. This greatly reduces the amount of storage space required for execution of the program.

Third, it tests to see if a given subject has both the criterion score and the item score for a particular correlation; only if he has both are his scores entered into the computation of the correlation coefficient.

It might be noted that the program allows the printing out of both the correlation and the covariance, the latter sometimes being useful in the interpretation of items.


```

C
C      *****CORRELATION-COVARIANCE PROGRAM*****
C
      INTEGER      SITEM(400),SCRIT(5),AITEM(5,400),AITMS(5,400),
C      ACRTS(5),ACRPR(5,400),ACRIT(5)
      DIMENSION    CRMN(5),CRSD(5),XCRPR(5),RITCR(5),XSD(5),FOR1(20),
C      FOR2(20),FOR3(20),IKEY(5),T(5),NS(5),ID(18),NSUB(5)
C
      IDATA=10
      IREAD=1
      IPRNT=3
1     KNS=0
      READ(IREAD,2) NCRT,NITM,ISTOP,ID,FOR1,FOR2,FOR3
2     FORMAT(2I3,I2,18A4/20A4/20A4/20A4)
C
      DO 5 I=1,NCRT
      NS(I)=0
      ACRIT(I)=0
      ACRTS(I)=0
      DO 5 J=1,NITM
      AITEM(I,J)=0
      AITMS(I,J)=0
5     ACRPR(I,J)=0
C
6     READ(IDATA,FOR1,END=10) (IKEY(I),I=1,NCRT),ITEST
      READ(IDATA,FOR2,ERR=98) (SCRIT(I),I=1,NCRT)
      IF(ITEST.EQ.0) GO TO 6
      READ(IDATA,FOR3) (SITEM(I),I=1,NITM)
C
      DO 9 I=1,NCRT
      IF(IKEY(I)) 7,9,7
7     NS(I)=NS(I)+1
      ACRIT(I)=ACRIT(I)+SCRIT(I)
      ACRTS(I)=ACRTS(I)+SCRIT(I)*SCRIT(I)
      DO 8 J=1,NITM
      AITEM(I,J)=AITEM(I,J)+SITEM(J)
      AITMS(I,J)=AITMS(I,J)+SITEM(J)*SITEM(J)

```

```

      ACRPR(I,J)=ACRPR(I,J)+SCRIT(I)*SITEM(J)
8  CONTINUE
9  CONTINUE
      KNS=KNS+1
      GO TO 6
C
10  REWIND IDATA
      DO 15 I=1,NCRT
          NSUB(I)=NS(I)
          IF(NS(I)) 11,12,11
11  T(I)=FLOAT(NS(I))
          GO TO 13
12  T(I)=.0001
13  SUM=FLOAT(ACRIT(I))
          SSUM=FLOAT(ACRTS(I))
          CRMN(I)=SUM/T(I)
          VAR=SSUM/T(I)
          CRSD(I)=(VAR-CRMN(I)*CRMN(I))
          IF(CRSD(I).LE.0.)GO TO 14
          CRSD(I)=SQRT(CRSD(I))
          GO TO 15
14  CRSD(I)=9999999999.
15  CONTINUE
C
      WRITE(IPRNT,3) ID,NCRT,NITM,(CRMN(I),I=1,NCRT)
3  FORMAT('1      ',18A4/10X,'CRITERIA =',I3,5X,'ITEMS =',I4/// ' ',
C      22X,'V1',20X,'V2',20X,'V3',20X,'V4',20X,'V5'//
C  ' MEAN',5(12X,F10.3))
      WRITE(IPRNT,17) (CRSD(I),I=1,NCRT)
17  FORMAT('  SD ',5(7X,F15.3))
      WRITE(IPRNT,18) (NSUB(I),I=1,NCRT)
18  FORMAT('    N ',12X,I4,4(18X,I4))
      WRITE(IPRNT,19)
19  FORMAT('0ITEM',6X,5(' R      SD      COV      '))//)
C
      DO 20 J=1,NITM
      DO 23 I=1,NCRT

```

```

SUM=FLOAT(AITEM(I,J))
SSUM=FLOAT(AITMS(I,J))
XMN=SUM/T(I)
XSD(I)=SSUM/T(I)-XMN*XMN
IF(XSD(I).LE.0) GO TO 21
XSD(I)=SQRT(XSD(I))
GO TO 22
21 XSD(I)=9999999999.
22 XCRP=FLOAT(ACRPR(I,J))
XCRPR(I)=XCRP/T(I)-XMN*CRMN(I)
RITCR(I)=XCRPR(I)/(XSD(I)*CRSD(I))
IF(ABS(RITCR(I)).LE.0.1) RITCR(I)=9999999999.
23 CONTINUE
WRITE(IPRNT,30)J, (RITCR(I),XSD(I),XCRPR(I) ,I=1,NCRT)
30 FORMAT(I4,6X,5(F5.2, F7.2,1X,F8.1,1X))
20 CONTINUE
C
98 CONTINUE
WRITE(IPRNT,40) KNS
40 FORMAT('0 ***END OF JOB*** NO. OF SUBJECTS =',I5)
IF(ISTOP.NE.0) GO TO 1
STOP
END

```

APPENDIX C: SUPPLEMENTARY TABLES

Table 20. Composition of MSAT factor scales

Factor	Reading	Same- Opposites	Analogies	Total
I	1,33,34 ^a	25,56,71	27,30,31,42, 45,47,58,60, 61,63,73-78	22
II	4,64-69	8		8
III	17,18,22	10	11,12,15,16, 28,29,43,46	12
IV	2,3,5,6,7, 19,23,35,36, 37,48,49,51, 53,54	9,10,24,26, 39,40,41,70		<u>24</u> 66

^aItem number.

Table 21. Composition of ENGL factor scales

Factor	Effectiveness	Mechanics	Total
I	30 ^a	32,33,38,41,48,53,56,60 65,67,68,69,72,75,76,79, 83,85,87,88	22
II	2,3,4-14,16, 18,20-25,28,29	31,45,47,59	27
III		34,35,36,37,46,52,55,61, 70,73,80,81,84,86	14
IV	17	42,43,50,54,58,62,64,66, 74,78,82	<u>12</u> 75

^aItem number.

Table 22. Composition of MATH factor scales

Factor	Items	Total
I	1,3,4,5,6,7,11,19,21, 23,27,28,30,31,32,34	16
II	41,50,55-65	13
III	2,9,10,12,13,14,15,18,29	9
IV	26,40,42,44,45,47,48,49,52	<u>9</u>
		47